

## Transcripts from a single full-length cDNA clone of hepatitis C virus are infectious when directly transfected into the liver of a chimpanzee

MASAYUKI YANAGI, ROBERT H. PURCELL, SUZANNE U. EMERSON, AND JENS BUKH\*

Hepatitis Viruses Section, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892

Contributed by Robert H. Purcell, June 9, 1997

**ABSTRACT** We have succeeded in constructing a stable full-length cDNA clone of strain H77 (genotype 1a) of hepatitis C virus (HCV). We devised a cassette vector with fixed 5' and 3' terminal and constructed multiple full-length cDNA clones of H77 in a single step by cloning of the entire ORF, which was amplified by long reverse transcriptase-PCR, directly into this vector. The infectivity of two complete full-length cDNA clones was tested by the direct intrabepatic injection of a chimpanzee with RNA transcripts. However, we found no evidence for HCV replication. Sequence analysis of these and 16 additional full-length clones revealed that seven clones were defective for polypeptide synthesis, and the remaining nine clones had 6–28 amino acid mutations in the predicted polypeptide compared with the consensus sequence of H77. Next, we constructed a consensus chimera from four of the full-length cDNA clones with just two ligation steps. Injection of RNA transcripts from this consensus clone into the liver of a chimpanzee resulted in viral replication. The sequence of the virus recovered from the chimpanzee was identical to that of the injected RNA transcripts. This stable infectious molecular clone should be an important tool for developing a better understanding of the molecular biology and pathogenesis of HCV.

Hepatitis C virus (HCV) is the most important cause of transfusion-associated and community-acquired non-A, non-B hepatitis (1, 2). The infection is characterized by a high rate of chronicity (>80%) (1), and it is estimated that about 4 million people in the United States and more than 100 million people worldwide are chronically infected with HCV (2). These individuals have a high risk of developing chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma (1). Thus, HCV is a major cause of morbidity and mortality worldwide, and there is an urgent need for prevention of infection and for effective treatment. However, there currently is no vaccine for HCV, and therapy with interferon, which is the only antiviral drug with proven efficacy for treatment of chronic hepatitis C, is effective in only 20–30% of patients (3).

HCV has been tentatively classified in a separate genus (*Hepacivirus*) of the *Flaviviridae* family. Like members of the genera of *Flavivirus* and *Pesivirus*, HCV has a positive-sense single-stranded RNA genome that is about 9.5 kb in length (4). It consists of a highly conserved 5' untranslated region (UTR) of approximately 340 nt, a single long ORF of approximately 9,000 nt, and a unique 3' UTR of 200–300 nt. The 3' UTR contains three distinct regions consisting of a short variable sequence (variable region) adjoining a poly U-UC region of variable length followed by a highly conserved terminal sequence (conserved region) of approximately 100 nt (5–8). The ORF of HCV encodes a large polypeptide precursor that is cotranslationally and post-

translationally cleaved into several structural and nonstructural proteins (9).

The genome of HCV is highly heterogeneous, and many genetic groups have been described (10). The most divergent HCV isolates differ from each other by more than 30% over the entire genome. Moreover, HCV circulates as a quasispecies of closely related genomes in a single infected individual.

At present the chimpanzee represents the only animal model for HCV infection (11, 12). However, the availability of this animal model is very limited. Furthermore, HCV replication is not efficient in cultured cells *in vitro* (13). The difficulties in propagating HCV have hindered basic research and the development of antiviral therapies and vaccines.

The genome of positive-strand RNA viruses functions as mRNA from which all viral proteins necessary for virus propagation are translated. Thus, genomic RNA, as well as RNA transcripts from full-length cDNA clones, should be infectious. In the *Flaviviridae* family, infectious transcripts of full-length cDNAs have been described for flaviviruses (14–18) and pestiviruses (19–23). An infectious molecular clone of HCV would be an important tool for better understanding of its molecular biology and pathogenesis. Given the extensive genotype diversity of HCV, it also may be important to have infectious clones representative of more than one genotype or variant. Therefore, it is important to devise a strategy for efficiently constructing such clones. Recently, by using the long reverse transcriptase-PCR (RT-PCR) method we succeeded in the synthesis of 7.5-kb DNA amplicons of hepatitis A virus from which infectious RNA was transcribed (24). We further demonstrated that it was possible to amplify the near full-length HCV genome in a single round of long RT-PCR (25). In the present study, we used this methodology to perform a detailed sequence analysis of the H77 strain of HCV. The sequence data were used to construct a cassette vector into which products from long RT-PCR were efficiently inserted to provide full-length genomic cDNA clones that were tested for their ability to yield infectious RNA transcripts.

### MATERIALS AND METHODS

**Virus Stock.** Plasma containing strain H77 of HCV was obtained from a patient in the acute phase of transfusion-associated non-A, non-B hepatitis (11). Strain H77 belongs to genotype 1a of HCV (26, 27). The consensus sequence for most of its genome has been determined (7, 26–28).

**Abbreviations:** HCV, hepatitis C virus; UTR, untranslated region; RT-PCR, reverse transcriptase-PCR; GE, genome equivalents; p.i., postinoculation.

**Data deposition:** The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF011751–AF011753).

\*To whom reprint requests should be addressed at: National Institutes of Health, National Institute of Allergy and Infectious Diseases, Laboratory of Infectious Diseases, Hepatitis Viruses Section, Building 7, Room 201, 7 Center Drive, MSC 0740, Bethesda, MD 20892-0740.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1997 by The National Academy of Sciences 0875-9424/97/948738-06\$02.00/0 PNAS is available online at <http://www.pnas.org>.

**RNA Purification.** Total RNA from 10  $\mu$ l of H77 plasma was extracted with the TRIzol system (GIBCO/BRL). The RNA pellet was resuspended in 100  $\mu$ l of 10 mM DTT with 5% (vol/vol) RNasin (20–40 units/ $\mu$ l) (Promega), and 10- $\mu$ l aliquots were stored at  $-80^{\circ}\text{C}$ . In subsequent experiments RT-PCR was performed on RNA equivalent to 1  $\mu$ l of H77 plasma, which contained an estimated  $10^4$  genome equivalents (GE) of HCV (29).

**Primers.** Primers were deduced from the genomic sequences of strain H77 previously published (see above) or determined in this study (Table 1). Primers for long RT-PCR were size-purified.

**cDNA Synthesis.** The RNA was denatured at  $65^{\circ}\text{C}$  for 2 min, and cDNA synthesis was performed in a 20- $\mu$ l reaction volume with Superscript II reverse transcriptase (GIBCO/BRL) at  $42^{\circ}\text{C}$  for 1 hr using specific antisense primers as described previously (24). The cDNA mixture was treated with RNase H and RNase T1 (GIBCO/BRL) for 20 min at  $37^{\circ}\text{C}$ .

**Amplification and Cloning of the 3' UTR.** The 3' UTR of strain H77 was amplified by PCR in two different assays. In both of these nested PCR reactions the first round of PCR was performed in a total volume of 50  $\mu$ l in 1 $\times$  buffer, 250  $\mu$ M each each deoxynucleoside triphosphate (Pharmacia), 20 pmol each external sense and antisense primers, 1  $\mu$ l of the Advantage KlenTaq polymerase mix (CLONTECH), and 2  $\mu$ l of the final cDNA reaction mixture. In the second round of PCR, 5  $\mu$ l of the first-round PCR mixture was added to 45  $\mu$ l of PCR mixture prepared as described above. Each round of PCR (35 cycles), which was performed in a DNA thermal cycler 480 (Perkin-Elmer), consisted of denaturation at  $94^{\circ}\text{C}$  for 1 min (in first cycle 1 min 30 sec), annealing at  $60^{\circ}\text{C}$  for 1 min, and elongation at  $68^{\circ}\text{C}$  for 2 min. In one experiment a region from NSSB to the conserved region of the 3' UTR was amplified with the external primers H9261F and H3'X58R, and the internal primers H9282F and H3'X45R (Table 1). In another experiment, a region from the variable region to the very end of the 3' UTR was amplified with the external primers H9375F and H3'X-35R, and the internal primers H9386F and H3'X-38R (Table 1, Fig. 1). Amplified products were purified with QIAquick PCR purification kit (Qiagen), digested with *Hind*III and *Xba*I (Promega), purified by either gel electrophoresis or phenol/chloroform extraction, and then cloned into the multiple cloning site of pGEM-9zf(–) (Promega) or pUC19 (Pharmacia). Cloning of cDNA into the vector was performed with T4 DNA ligase (Promega) by standard procedures.

**Amplification of Near Full-Length H77 Genomes by Long PCR.** The reactions were performed in a total volume of 50  $\mu$ l in 1 $\times$  buffer, 250  $\mu$ M each deoxynucleoside triphosphate, 10 pmol of sense and antisense primers each, 1  $\mu$ l of the Advantage KlenTaq polymerase mix, and 2  $\mu$ l of the cDNA reaction mixture (24, 25). A single PCR round of 35 cycles was performed in a Robocycler thermal cycler (Stratagene) and consisted of denaturation at  $99^{\circ}\text{C}$  for 35 sec, annealing at  $67^{\circ}\text{C}$  for 30 sec, and

elongation at  $68^{\circ}\text{C}$  for 10 min during the first five cycles, 11 min during the next 10 cycles, 12 min during the following 10 cycles, and 13 min during the last 10 cycles. To amplify the complete ORF of HCV by long RT-PCR we used the sense primers H1 or A1 deduced from the 5' UTR and the antisense primer H9417R deduced from the variable region of the 3' UTR (Table 1, Fig. 1).

**Construction of Full-Length H77 cDNA Clones.** The long PCR products amplified with H1 and H9417R primers were cloned directly into pGEM-9zf(–) after digestion with *Not*I and *Xba*I (Promega) (Fig. 1). We obtained only two clones with inserts of the expected size, pH21, and pH50. Next, the chosen 3' UTR was cloned into both pH21, and pH50, after digestion with *Afl*III and *Xba*I (New England Biolabs). DH5 $\alpha$  competent cells (GIBCO/BRL) were transformed and selected with Luria-Bertani agar plates containing 100  $\mu$ g/ml ampicillin (Sigma). Then the selected colonies were cultured in Luria-Bertani liquid containing ampicillin at  $30^{\circ}\text{C}$  for 18–20 hr (transformants containing full-length or near full-length cDNA of H77 produced a very low yield of plasmid when cultured at  $37^{\circ}\text{C}$  or for more than 24 hr). After small-scale preparation (Wizard Plus Minipreps DNA Purification Systems, Promega) each plasmid was retransformed to select a single clone, and large-scale preparation of plasmid DNA was performed with a Qiagen plasmid Maxi kit.

**Cloning of Long RT-PCR Products into a Cassette Vector.** To improve the efficiency of cloning, we constructed a vector with consensus 5' and 3' termini of HCV strain H77 (Fig. 1). This cassette vector (pCV) was obtained by cutting out the *Bam*HI fragment (nucleotides 1,358–7,530 of the H77 genome) from pH50, followed by religation. Next, the long PCR products of H77 amplified with H1 and H9417R or A1 and H9417R primers were purified (GeneClean spin kit; BIO 101) and cloned into pCV after digestion with *Age*I and *Afl*III (New England Biolabs) or with *Pvu*AI (isoschizomer of *Age*I) and *Bst*PI (isoschizomer of *Afl*III) (Boehringer Mannheim). Large-scale preparations of the plasmids containing full-length cDNA of H77 were performed as described above.

**Construction of H77 Consensus Chimeric cDNA Clone.** To construct a full-length cDNA clone of H77 with an ORF encoding the consensus amino acid sequence, we made a chimera from four of the cDNA clones obtained above. This consensus chimera, pCV-H77C, was constructed in two ligation steps by using standard molecular procedures and convenient cleavage sites and involved first a two-piece ligation and then a three-piece ligation. Large-scale preparation of pCV-H77C was performed as described above.

**In Vitro Transcription.** Plasmids containing the full-length HCV cDNA were linearized with *Xba*I (Promega) and purified by phenol/chloroform extraction and ethanol precipitation. A 100- $\mu$ l reaction mixture containing 10  $\mu$ g of linearized plasmid DNA, 1 $\times$  transcription buffer, 1 mM ATP, CTP, GTP, and UTP, 10 mM DTT, 4% (vol/vol) RNasin (20–40 units/ $\mu$ l), and 2  $\mu$ l of T7 RNA polymerase (Promega) was incubated at  $37^{\circ}\text{C}$  for 2 hr.

Table 1. Oligonucleotides used for PCR amplification of strain H77 of HCV

Designation	Sequence (5' $\rightarrow$ 3')*
H9261F	GGCTACAGCGGGGAGACATTTATCACAGC
H3'X58R	TCAATGCGGCTCACGACCTTTTCACAGCTAG
H9282F	<b>GTCCAGCTTATCACAGCGTGTCTCATGCCCGCCCG</b>
H3'X45R	<b>CGTCTCTAGAGGACCTTTTCACAGCTAGCCGTGACTAGGG</b>
H9375F	TGAAGGTTGGGGTAAACACTCCGGCCCTCTTAGGCCATT
H3'X-35R	ACATGATCTGCAGAGAGGCCAGTATCAGCACTCTC
H9386F	<b>GTCCAGCTTACCGGTAAACACTCCGGCCCTCTTAGGCCATTTCCTG</b>
H3'X-38R	<b>CGTCTCTAGACATGATCTGCAGAGAGGCCAGTATCAGCACTCTCTGC</b>
H1	<b>TTTTTTTTCGCGCCGCTAAATACGATCACTATAGCCAGCCCTGATGGGGCGGACACTCCACCATG</b>
A1	ACTGTCTTCACGCAAGAGCGTCTAGCCAT
H9417R	<b>CGTCTCTAGACAGGAAATGGCTTAGAGGCCGGAATGTTTACC</b>

\*HCV sequences are shown in lightface text, non-HCV-specific sequences are shown in boldface, and artificial cleavage sites used for cDNA cloning are underlined. The core sequence of the T7 promoter in primer H1 is shown in italics.



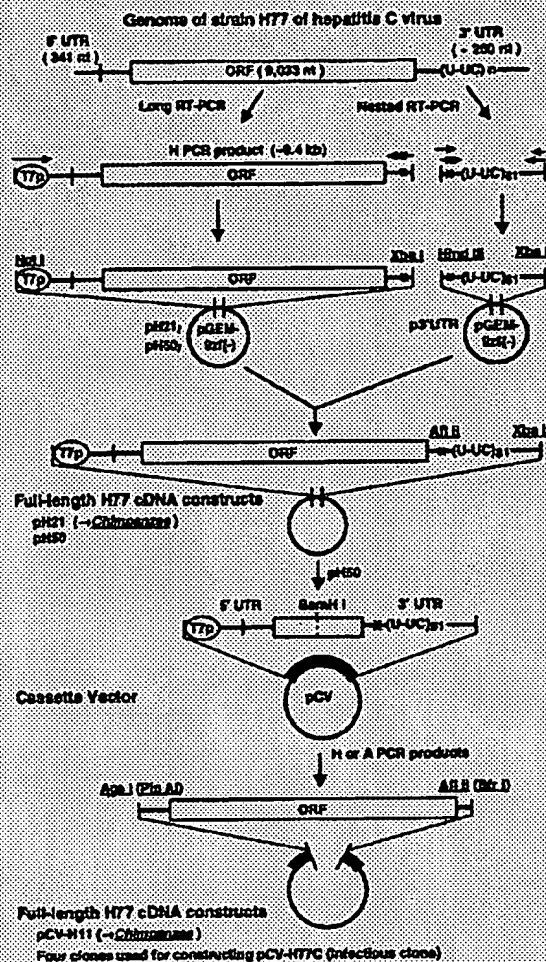


FIG. 1. Strategy for the construction of full-length cDNA clones of HCV strain H77. The long PCR products amplified with H1 and H9417R primers were cloned directly into pGEM-9zf(+) after digestion with *Nde*I and *Xba*I (pH21; and pH50). Next, the 3' UTR was cloned into both pH21 and pH50 after digestion with *A*III and *Xba*I (pH21 and pH50). pH21 was tested for infectivity in a chimpanzee. To improve the efficiency of cloning, we constructed a cassette vector with consensus 5' and 3' termini of H77. This cassette vector (pCV) was obtained by cutting out the *Bam*HI fragment (nucleotides 1,358–7,530 of the H77 genome) from pH50, followed by religation. Finally, the long PCR products of H77 amplified with primers H1 and H9417R (H product) or primers A1 and H9417R (A product) were cloned into pCV after digestion with *A*geI and *A*III or with *P*inAI and *B*/H. The latter procedure yielded multiple complete cDNA clones of strain H77 of HCV.

Five microliters of the reaction mixture was analyzed by agarose gel electrophoresis followed by ethidium bromide staining. The transcription reaction mixture was diluted with 400  $\mu$ l of ice-cold PBS without calcium or magnesium, immediately frozen on dry ice and stored at  $-80^{\circ}\text{C}$ . The final nucleic acid mixture was injected into chimpanzees within 24 hr.

**Intrahepatic Transfection of Chimpanzees.** Laparotomy was performed, and aliquots from two transcription reactions were injected into six sites of the exposed liver (30). Serum samples were collected weekly from chimpanzees and monitored for liver enzyme levels and anti-HCV antibodies. Weekly samples of 100  $\mu$ l of serum were tested for HCV RNA in a highly sensitive nested RT-PCR assay with AmpliTaq Gold (Perkin-Elmer) (29, 31). The genome titer of HCV was

estimated by testing 10-fold serial dilutions of the extracted RNA in the RT-PCR assay (29). The two chimpanzees used in this study were maintained under conditions that met all requirements for their use in an approved facility.

The consensus sequence of the complete ORF from HCV genomes recovered at week 2 postinoculation (p.i.) was determined by direct sequencing of PCR products obtained in long RT-PCR with primers A1 and H9417R followed by nested PCR of 10 overlapping fragments. The consensus sequence of the variable region of the 3' UTR was determined by direct sequencing of an amplicon obtained in nested RT-PCR as described above. Finally, we amplified selected regions independently by nested RT-PCR with AmpliTaq Gold.

**Sequence Analysis.** Both strands of DNA from PCR products, as well as plasmids, were sequenced with the Applied Biosystems PRISM Dye Termination Cycle Sequencing Ready Reaction Kit using AmpliTaq DNA polymerase (Perkin-Elmer) and about 100 specific sense and antisense sequence primers.

We determined the consensus sequence of HCV strain H77 in two different ways. In one approach we directly sequenced overlapping PCR products amplified in nested RT-PCR from the H77 plasma sample. The sequence analyzed (nucleotides 35–9,417) included the entire genome except the very 5' and 3' termini. In the second approach, the consensus sequence of nucleotides 157–9,384 was deduced from the sequences of 18 full-length cDNA clones.

## RESULTS

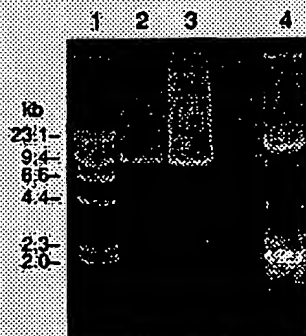
**Variability in the Sequence of the 3' UTR of HCV Strain H77.** The heterogeneity of the 3' UTR was analyzed by cloning and sequencing of DNA amplicons obtained in nested RT-PCR. We analyzed 19 clones, which included sequences of the entire variable region, the poly U-UC region, and the adjacent 19 nt of the conserved region, and 65 clones containing sequences of the entire poly U-UC region and the first 63 nts of the conserved region. We confirmed that the variable region consisted of 43 nts, including two conserved termination codons (32). The sequence of the variable region was highly conserved within H77 because only three point mutations were found among the 19 clones analyzed. A poly U-UC region was present in all 84 clones analyzed. However, its length varied from 71–141 nts. The length of the poly U region was 9–103 nts, and that of the poly UC region was 35–85 nts. The number of C residues increased toward the 3' end of the poly UC region but the sequence of this region was not conserved. The first 63 nts of the conserved region were highly conserved among the clones analyzed, with a total of only 14 point mutations. To confirm the validity of the analysis, we reamplified the 3' UTR directly from a full-length cDNA clone of HCV (see below) by the nested-PCR procedure with the primers in the variable region and at the very 3' end of the HCV genome and cloned the PCR product. Eight clones had 1–7 nt deletions in the poly U region. Furthermore, although the C residues of the poly UC region were maintained, the spacing of these varied because of 1–2 nt deletions of U residues. These deletions must be artifacts introduced by PCR, and such mistakes may have contributed to the heterogeneity originally observed in this region. However, the conserved region of the 3' UTR was amplified correctly, suggesting that the deletions were due to difficulties in transcribing a highly repetitive sequence.

One of the 3' UTR clones was selected for engineering of full-length cDNA clones of H77. This clone had the consensus variable sequence except for a single point mutation introduced to create an *A*III cleavage site, a poly U-UC stretch of 81 nts with the most commonly observed UC pattern and the consensus sequence of the complete conserved region of 101 nt, including the distal 38 nt that originated from the antisense primer used in the amplification. After linearization with *Xba*I, the DNA template of this clone had the authentic 3' end.

**The Entire ORF of H77 Amplified in One Round of Long RT-PCR.** We previously demonstrated that a 9.25-kb fragment of the HCV genome from the 5' UTR to the 3' end of NSSB could be amplified from 10<sup>4</sup> GE of H77 by a single round of long RT-PCR (25). In the current study, by optimizing primers and cycling conditions, we were able to amplify the entire ORF of H77 in a single round of long RT-PCR with primers from the 5' UTR and the variable region of the 3' UTR. In fact, we could amplify 9.4 kb of the H77 genome (H product: from the very 5' end to the variable region of the 3' UTR) from 10<sup>5</sup> GE or 9.3 kb (A product: from within the 5' UTR to the variable region of the 3' UTR) from 10<sup>4</sup> GE or 10<sup>3</sup> GE, in a single round of long RT-PCR (Fig. 2). The PCR products amplified from 10<sup>5</sup> GE of H77 were used for engineering full-length cDNA clones (see below).

**Construction of Multiple Full-Length cDNA Clones of H77 in a Single Step by Cloning of Long RT-PCR Amplicons Directly into a Cassette Vector with Fixed 5' and 3' Termini.** We first attempted direct cloning of the long PCR products (H), which contained a 5' T7 promoter, the authentic 5' end, the entire ORF of H77 and a short region of the 3' UTR, into pGEM-2d(-) vector by *NorI* and *XbaI* digestion. However, among the 70 clones examined all but two had inserts that were shorter than predicted. Sequence analysis identified a second *NorI* site in the majority of clones, which resulted in deletion of the nucleotides past position 9,221. Only two clones (pH21 and pH50) were missing the second *NorI* site and had the expected 5' and 3' sequences of the PCR product. Therefore, full-length cDNA clones (pH21 and pH50) were constructed by inserting the chosen 3' UTR into pH21 and pH50, respectively. Sequence analysis revealed that clone pH21 had a complete full-length sequence of H77; this clone was tested for infectivity. The second clone, pH50, had 1 nt deletion in the ORF at position 6,365; this clone was used to make a cassette vector.

Because we could not amplify the entire genome in one step but we could amplify the complete ORF, we constructed a cassette vector with fixed 5' and 3' termini as an intermediate of full-length cDNA clones. We constructed this vector (pCV) by digestion of clone pH50 with *Bam*HI, followed by religation. This shortened the plasmid so that it could be readily distinguished from plasmids containing the full-length insert. Attempts to clone long RT-PCR products (H) into pCV by *Age*I and *Afl*II yielded only 1 of 23 clones with an insert of the expected size. To increase the efficiency of cloning, we re-



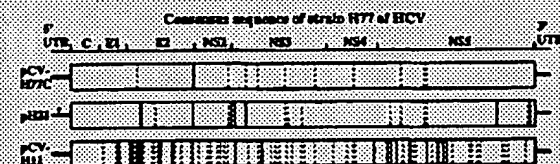
**FIG. 2.** Gel electrophoresis of long RT-PCR amplicons of the entire ORF of H77 and the transcription mixture of the infectious clone of H77. The complete ORF was amplified by long RT-PCR with the primers H1 or A1 and H9417R from 10<sup>5</sup> GE of H77. A total of 10  $\mu$ g of the consensus chimeric clone (pCV-H77C) linearized with *Xba*I was transcribed in a 100- $\mu$ l reaction with T7 RNA polymerase. Five microliters of the transcription mixture was analyzed by gel electrophoresis and the remainder of the mixture was injected into a chimpanzee. Lane 1, molecular weight marker; lane 2, products amplified with primers H1 and H9417R; lane 3, products amplified with primers A1 and H9417R; lane 4, transcription mixture containing the RNA transcripts and linearized clone pCV-H77C (12.5 kb).

peated the procedure but used *P*inAI and *B*frI instead of the respective isoschizomers *A*geI and *A*flII. By this protocol, 24 of 31 H clones and 30 of 35 A clones had the full-length cDNA of H77 as evaluated by restriction enzyme digestion. A total of 16 clones, selected at random, were each retransformed, and individual plasmids were purified and completely sequenced.

**Two Full-Length cDNA Clones of H77 Were Not Infectious in a Chimpanzee.** The infectivity of the two first complete full-length cDNA clones of H77 (pH21 and pCV-H11) we obtained were tested by the intrahepatic injection of a chimpanzee with RNA transcripts. The transcription mixture from each clone was injected into three sites of the exposed liver. However, we found no evidence for HCV replication in the chimpanzee. Weekly serum samples were negative for HCV RNA at weeks 1–17 p.i. in a highly sensitive nested RT-PCR assay. The cDNA template injected along with the RNA transcripts also was not detected in this assay. The chimpanzee remained negative for antibodies to HCV throughout follow-up. Thus, clones pH21 and pCV-H11 were not infectious.

**Failure to Identify a Consensus Clone by Sequence Analysis of the ORF of 18 Full-Length cDNA Clones.** We performed sequence analysis of 18 full-length clones. Clone pH21, which was not infectious (see above), had seven aa substitutions in the entire predicted polyprotein compared with the consensus sequence of H77 (Fig. 3). The most notable mutation, at position 1,026, changed L to Q, which altered the cleavage site between NS2 and NS3 (33). Clone pCV-H11, which also was not infectious (see above), had 21 aa substitutions in the predicted polyprotein compared with the consensus sequence of H77 (Fig. 3). One aa mutation (position 564) eliminated a highly conserved C residue in the E2 protein (34). These two mutations might conceivably account for the lack of infectivity of the two clones but any of the other mutations might have been responsible. Alternatively, the 5' and 3' termini might not have been viable.

The infectivity of the remaining 16 full-length clones was not tested. However, four clones had a single nucleotide deletion in the ORF of H77, which would result in a frame shift, and an additional clone had two nucleotide insertions in the ORF, which also would cause a frame shift. Finally, stop codons were identified in the ORF of two more clones. Thus, these seven clones were defective for polyprotein synthesis and clearly would not be infectious. This left us with nine potentially infectious clones. However, these clones had 6–28 aa mutations in the predicted polyprotein compared with the consensus sequence of H77. More importantly, each clone had aa mutations not observed in other isolates of HCV (10).



**FIG. 3.** Diagram of the genome organization of HCV strain H77 and the genetic heterogeneity of individual full-length clones compared with the consensus sequence of H77. Solid lines represent amino acid changes. Dashed lines represent silent mutations. A \* in pH21 represents a point mutation at nucleotide 58 in the 5' UTR. In the ORF, the consensus chimeric clone pCV-H77C had 11 nucleotide differences (at positions 1,625 (C → T), 2,709 (T → C), 3,380 (A → G), 3,710 (C → T), 3,914 (G → A), 4,463 (T → C), 5,058 (C → T), 5,834 (C → T), 6,734 (T → C), 7,154 (C → T), and 7,202 (T → C)) and one amino acid change (F → L at amino acid 790) compared with the consensus sequence of H77. This clone was infectious. Clone pH21 and pCV-H11 had 19 nucleotide (7 amino acid) and 64 nucleotide (21 amino acid) differences, respectively, compared with the consensus sequence of H77. These two clones were not infectious. A single point mutation in the 3' UTR at nucleotide 9,406 (G → A) introduced to create an *Afl*II cleavage site is not shown.



Transcripts from a cDNA clone representing the consensus sequence of HCV strain H77 are infectious when transfected into the liver of a chimpanzee. It is well established that RNA transcripts of cDNA clones encompassing the consensus sequences of flaviviruses or pestiviruses are infectious (14, 19). Given the limited availability of chimpanzees, which currently represent the only reliable model for HCV propagation, we decided to test the infectivity of a consensus clone of H77. A consensus chimera was constructed from four of the full-length cDNA clones with just two ligation steps. The final construct, pCV-H77C, had 11 nucleotide differences from the consensus sequence of H77 in the ORF (Fig. 3). However, 10 of these nucleotide differences represented silent mutations. The chimeric clone differed from the consensus sequence at only one amino acid (L instead of F at position 790). Among the 18 ORFs analyzed above, the F residue was found in 11 clones and the L residue in seven clones. However, the L residue was dominant in other isolates of genotype 1a, including a first passage of H77 in a chimpanzee (27). Thus, we did not believe that this amino acid difference was critical for infectivity.

To test the infectivity of the consensus chimera clone of H77 we performed intrahepatic transfection of a chimpanzee. The pCV-H77C clone was linearized with *Xba*I and transcribed *in vitro* by T7 RNA polymerase (Fig. 2). The transcription mixture was next injected into six sites of the chimpanzee liver. We used exactly the same total amount of DNA template and transcription mixture as used in the unsuccessful transfection described above in which viral sequences were not detected at any time after injection. The chimpanzee became infected with HCV as measured by detection of  $10^2$  GE/ml of viral genome at week 1 p.i. Furthermore, the HCV titer increased to  $10^4$  GE/ml at week 2 p.i. and had reached  $10^6$  GE/ml by week 8 p.i.

We analyzed the sequence of the HCV genomes from the serum sample collected at week 2 p.i. The consensus sequence of nucleotides 298–9,375 of the recovered genomes was determined by direct sequencing of PCR products obtained in long RT-PCR followed by nested PCR of 10 overlapping fragments. The identity to clone pCV-H77C sequence was 100%. The consensus sequence of nucleotides 96–291, 1,328–1,848, 3,585–4,106, 4,763–5,113, and 9,322–9,445 was determined from PCR products obtained in different nested RT-PCR assays. The identity of these sequences with pCV-H77C was also 100%. These latter regions contained four mutations unique to the consensus chimera, including the artificial *A/H* cleavage site in the 3' UTR. Therefore, RNA transcripts of this clone of HCV were infectious.

## DISCUSSION

In the present study, we have constructed an infectious clone of strain H77 of HCV. This clone represents the consensus sequence of H77 and contains a 5' UTR of 341 nt, an ORF of 9,033 nt, and a 3' UTR of 225 nt (a variable region of 43 nt with a single point mutation, poly U-UC region of 81 nt, and a conserved region of 101 nt). Two other full-length cDNA clones of H77 with the same termini were not infectious, most likely due to one or more of the amino acid mutations found in the predicted polyprotein.

Due to the lack of reliable *in vitro* propagation systems of HCV we could not perform screening for infectivity in cell cultures. We previously established an *in vivo* transfection system for RNA transcripts of infectious clones of hepatitis A virus in tamarins (30), as well as in chimpanzees. Therefore, the infectivity of RNA transcripts of full-length HCV clones was tested by injecting transcription mixtures into the liver of chimpanzees.

We demonstrated that RNA transcripts of the consensus chimera of H77 were infectious *in vivo* and that the titer of genomic sequences recovered from the serum of the chimpanzee increased over time. Sequence analysis demonstrated that the recombinant virus recovered from the chimpanzee after the increase in viral titer had a sequence identical to that of the chimeric clone, including several unique mutations. It is possible that the cDNA injected with the transcribed RNA was itself

infectious. However, we previously demonstrated that the cDNA clone of hepatitis A virus, inserted into a similar vector also lacking eukaryotic promoters, was not infectious when injected into the liver of tamarins (30).

The viremic pattern observed in the early phase of the infection with the recombinant virus was similar to that observed in chimpanzees inoculated intravenously with strain H77 or other strains of HCV (ref. 12; J.B., unpublished data). Viral hepatitis normally develops after week 8 p.i. However, biological variation exists, and some chimpanzees develop only minimal evidence of hepatitis. Thus, a careful study of the phenotype of the infectious clone of HCV would require transfection of several chimpanzees.

Reports of infectious transcripts of full-length cDNAs have been published for several members of the genus *Flavivirus* (14–18) and more recently for members of the genus *Pestivirus* (19–23). Most of these infectious cDNA clones were constructed by screening of cDNA libraries followed by the assembly of clones representing the consensus sequence. These common procedures are complicated and time-consuming, particularly in the case of HCV, which exists as heterogeneous quasispecies. We previously showed that full-length infectious cDNA clones of hepatitis A virus could be engineered in one step by cloning of long RT-PCR amplicons (35). We had to modify this approach for HCV, because we were not able to amplify the entire genome by long RT-PCR, probably because of the strong predicted secondary structure, as well as the presence of a long poly U-UC region in the 3' UTR. Instead we used a cassette vector with fixed termini of the HCV genome and cloned the entire ORF, amplified in long RT-PCR, directly into this vector. Thus, with this strategy we were able to construct full-length HCV clones in two steps.

One conclusion of our study is that a high proportion of HCV genomes probably are defective. We found that two of the complete full-length clones constructed were not infectious. Sequence analysis revealed that 7 of 16 additional clones were defective for polyprotein synthesis, and all clones had multiple amino acid mutations compared with the consensus sequence of the parent strain. We cannot rule out that some of these mutations were introduced in the RT-PCR and cloning procedures. However, the infectivity titer of H77 is 1–2  $\log_{10}$  lower than the genome titer suggesting that less than 10% of the genomes are infectious (refs. 11 and 29; R.H.P., unpublished data). Moormann et al. (19) reported that five aa changes (compared with the consensus sequence) in the polyprotein accounted for noninfectivity of classical swine fever virus. Furthermore, at a recent meeting<sup>†</sup> it was reported by C. M. Rice and A. A. Kolykhalov that, of numerous clones of H77 they tested, only those engineered to contain the consensus sequence were infectious. Although nine of our clones potentially could be infectious they all had six or more mutations compared with the consensus sequence of strain H77. Because of the limited availability of chimpanzees, we constructed a consensus chimera clone of strain H77. By using four of the full-length clones we constructed this clone in just two steps.

The infectious nature of the consensus chimera indicated that the regions of the 5' and 3' UTRs incorporated into the cassette vector did not destroy viability and suggested that it indeed was one or more of the coding mutations in each of the original two constructs that were lethal. Therefore, it should be possible to use the cassette vector to construct infectious cDNA clones of other HCV strains but it probably will be necessary to insert the consensus sequence for each ORF tested.

The infectious clone of HCV generated in the present study contained the 5' and 3' termini previously published by others (7, 36). Because the core sequence of the T7 promoter was posi-

<sup>†</sup>Fourth International Meeting on Hepatitis C Virus and Related Viruses, March 6–10, 1997, Kyoto, Japan.

tioned immediately before the authentic G of the 5' UTR sequence, the RNA transcripts should include the exact 5' end of HCV. Furthermore, after plasmid linearization with *Xba*I, the template DNA and therefore the RNA transcripts should end exactly at the 3' end. The 5' UTR of HCV previously has been found to contain an internal ribosomal entry site (4). We found that uncapped RNA was infectious, indicating that cap-independent translation initiation is used by HCV for gene expression. Thus, the internal ribosomal entry site must be functional *in vivo*. It is unclear which elements of the 3' UTR are required for infectivity of HCV. We found that the variable region was highly conserved in H77. However, the point mutation introduced in this region for cloning purposes was tolerated in the infectious clone. The length and sequence of the poly U-UC region varied greatly in H77. Our infectious clone of H77 contained a poly U-UC of intermediate length. Finally, we found that the conserved region of the 3' UTR was highly conserved, and our infectious clone contained the consensus sequence (101 nts) of this region (7). In one previous study it was reported that the RNA transcripts from a cDNA clone of HCV-1, the prototype of genotype 1a, resulted in a low level of replication after transfection into a human hepatoma cell line (37). This cDNA clone did not contain the conserved region of the 3' UTR. Furthermore, the infectivity of this clone was not tested *in vivo*. Recently, it was reported that deletions in the variable region of the 3' UTR of Dengue virus and Kunjin virus did not render these viruses noninfectious (38, 39). However, the conserved core sequences of the 3' UTR were critical for viral replication. Further studies, which are now possible, are needed to determine which regions of the 3' UTR are essential for HCV replication.

Instability of plasmids containing the full-length cDNA constructs of flaviviruses or pestiviruses created major obstacles to the construction of infectious clones for these viruses (14–23). We used the high copy vector pGEM-2zf(–) and DH5 $\alpha$  cells to prepare full- or near full-length cDNA clones of HCV. We found that transformants produced a very low yield of plasmid DNA when cultured at 37°C or for more than 24 hr. However, we could obtain a good yield of plasmids when cultures were incubated at 30°C for 18–20 hr. Furthermore, the sequence of the chimeric clone of H77, after retransformation and large-scale preparation, was identical to the sequence of the inserts from the four clones used in the construction. These results indicate that our infectious clone was stable in DH5 $\alpha$  cells.

The quasispecies nature of HCV is believed to be important for viral persistence (10). The successful infection of a chimpanzee with a single clone of HCV provides us with a unique opportunity to study the quasispecies nature of this virus and its evolution *in vivo*, as well as its importance for viral persistence.

In conclusion, we have constructed a genetically stable infectious clone of HCV. The approach used to engineer this clone should be applicable to construction of full-length infectious cDNAs of other HCV strains (10), as well as of a number of recently discovered related viruses (see ref. 40). Furthermore, this infectious clone might be helpful in developing effective *in vitro* propagation systems. Finally, the availability of an infectious clone of HCV makes it possible to study in detail the mechanisms of viral replication and pathogenesis.

We thank Ms. L. Rasmussen and other staff members at Science Applications International Corporation, Frederick, MD, for assistance in sequence analysis and Mr. M. Shapiro and other staff members at Bioqual, Rockville, MD, for providing animal care. We are grateful to Ms. T. S. Tsareva for primer synthesis. We also thank Dr. R. Tellier for helpful discussions and Dr. H. J. Alter for kindly providing plasma of Patient H. This study was supported in part by National Institutes of Health Contracts N01-AI-52705, N01-AI-45180, and N01-CO-56000.

1. Alter, H. J. & Seeff, L. B. (1993) in *Viral Hepatitis: Scientific Basis and Clinical Management*, eds. Zuckerman, A. J. & Thomas, H. C. (Churchill Livingstone, Edinburgh), pp. 467–498.

2. Alter, M. J. (1995) *Semin. Liver Dis.* 15, 5–14.
3. Fried, M. W. & Hootnagle, J. H. (1995) *Semin. Liver Dis.* 15, 82–91.
4. Houghton, M. (1996) in *Fields Virology*, eds. Fields, B. N., Knipe, D. M. & Howley, P. M. (Lippincott-Raven, Philadelphia), 3rd Ed., pp. 1035–1058.
5. Tanaka, T., Kato, N., Cho, M.-J. & Shimotohno, K. (1995) *Biochem. Biophys. Res. Commun.* 215, 744–749.
6. Tanaka, T., Kato, N., Cho, M.-J., Sugiyama, K. & Shimotohno, K. (1996) *J. Virol.* 70, 3307–3312.
7. Kolythakov, A. A., Feinstone, S. M. & Rice, C. M. (1996) *J. Virol.* 70, 3363–3371.
8. Yamada, N., Tanihara, K., Takada, A., Yoritani, T., Tsutsumi, M., Shimomura, H., Tsuji, T. & Date, T. (1996) *Virology* 223, 255–261.
9. Rice, C. M. (1996) in *Fields Virology*, eds. Fields, B. N., Knipe, D. M. & Howley, P. M. (Lippincott-Raven, Philadelphia), 3rd Ed., pp. 931–959.
10. Bukh, J., Miller, R. H. & Purcell, R. H. (1995) *Semin. Liver Dis.* 15, 41–63.
11. Feinstone, S. M., Alter, H. J., Dienes, H. P., Shimizu, Y., Popper, H., Blackmore, D., Sly, D., London, W. T. & Purcell, R. H. (1981) *J. Infect. Dis.* 144, 588–598.
12. Shimizu, Y. K., Weiner, A. J., Rosenblatt, J., Wong, D. C., Shapiro, M., Popkin, T., Houghton, M., Alter, H. J. & Purcell, R. H. (1990) *Proc. Natl. Acad. Sci. USA* 87, 6441–6444.
13. Shimizu, Y. K., Iwamoto, A., Hijikata, M., Purcell, R. H. & Yoshikura, H. (1992) *Proc. Natl. Acad. Sci. USA* 89, 5477–5481.
14. Rice, C. M., Grakoui, A., Galler, R. & Chambers, T. J. (1989) *New Biol.* 1, 285–296.
15. Lai, C.-J., Zhao, B., Hori, H. & Bray, M. (1991) *Proc. Natl. Acad. Sci. USA* 88, 5139–5143.
16. Sumiyoshi, H., Hoke, C. H. & Trent, D. W. (1992) *J. Virol.* 66, 5425–5431.
17. Khromykh, A. A. & Westaway, E. G. (1994) *J. Virol.* 68, 4580–4588.
18. Mandl, C. W., Ecker, M., Holzmann, H., Kunz, C. & Heinz, F. X. (1997) *J. Gen. Virol.* 78, 1049–1057.
19. Moormann, R. J. M., van Gennip, H. G. P., Miedema, G. K. W., Huik, M. M. & van Rijn, P. A. (1996) *J. Virol.* 70, 763–770.
20. Meyers, G., Thiel, H.-J. & Rumenapf, T. (1996) *J. Virol.* 70, 1588–1595.
21. Ruggli, N., Tratschin, J.-D., Mittelholzer, C. & Hofmann, M. A. (1996) *J. Virol.* 70, 3478–3487.
22. Meyers, G., Tautz, N., Becker, P., Thiel, H.-J. & Kunz, B. M. (1996) *J. Virol.* 70, 8606–8613.
23. Vassilev, V. B., Collett, M. S. & Donis, R. O. (1997) *J. Virol.* 71, 471–478.
24. Tellier, R., Bukh, J., Emerson, S. U. & Purcell, R. H. (1996) *Proc. Natl. Acad. Sci. USA* 93, 4370–4373.
25. Tellier, R., Bukh, J., Emerson, S. U., Miller, R. H. & Purcell, R. H. (1996) *J. Clin. Microbiol.* 34, 3085–3091.
26. Ogata, N., Alter, H. J., Miller, R. H. & Purcell, R. H. (1991) *Proc. Natl. Acad. Sci. USA* 88, 3392–3396.
27. Inchausti, G., Zebec, S., Lee, D.-H., Sugitani, M., Nasoff, M. & Prince, A. M. (1991) *Proc. Natl. Acad. Sci. USA* 88, 10292–10296.
28. Farci, P., Shimoda, A., Wong, D., Cabezon, T., De Giannis, D., Strazzera, A., Shimizu, Y., Shapiro, M., Alter, H. J. & Purcell, R. H. (1996) *Proc. Natl. Acad. Sci. USA* 93, 15394–15399.
29. Yanagi, M., Bukh, J., Emerson, S. U. & Purcell, R. H. (1996) *J. Infect. Dis.* 174, 1324–1327.
30. Emerson, S. U., Lewis, M., Govindarajan, S., Shapiro, M., Moskal, T. & Purcell, R. H. (1992) *J. Virol.* 66, 6649–6654.
31. Bukh, J., Purcell, R. H. & Miller, R. H. (1992) *Proc. Natl. Acad. Sci. USA* 89, 187–191.
32. Han, J. H. & Houghton, M. (1992) *Nucleic Acids Res.* 20, 3520.
33. Reed, K. E., Grakoui, A. & Rice, C. M. (1995) *J. Virol.* 69, 4127–4136.
34. Okamoto, H., Kurai, K., Okada, S.-I., Yamamoto, K., Iizuka, H., Tanaka, T., Fukuda, S., Tsuda, F. & Mishiro, S. (1992) *Virology* 188, 331–341.
35. Tellier, R., Bukh, J., Emerson, S. U. & Purcell, R. H. (1997) *Proceedings of the 1996 International Symposium on Viral Hepatitis and Liver Disease*, in press.
36. Han, J. H., Shyamala, V., Richman, K. H., Brauer, M. J., Irvine, B., Urdea, M. S., Tekamp-Olson, P., Kuo, G., Choo, Q.-L. & Houghton, M. (1991) *Proc. Natl. Acad. Sci. USA* 88, 1711–1715.
37. Yoo, B. J., Selby, M. J., Choe, J., Suh, B. S., Choi, S. H., Joh, J. S., Nuovo, G. J., Lee, H.-S., Houghton, M. & Han, J. H. (1995) *J. Virol.* 69, 32–38.
38. Men, R., Bray, M., Clark, D., Chanock, R. M. & Lai, C.-J. (1996) *J. Virol.* 70, 3930–3937.
39. Khromykh, A. A. & Westaway, E. G. (1997) *J. Virol.* 71, 1497–1505.
40. Bukh, J. & Apper, C. L. (1997) *Virology* 229, 429–436.